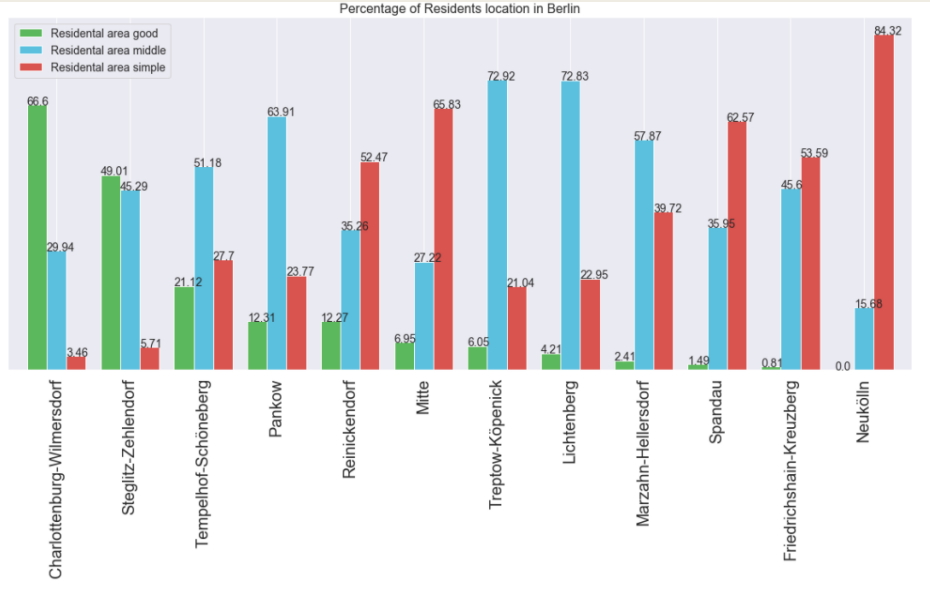
THE BATTLE OF NEIGHBORHOODS-

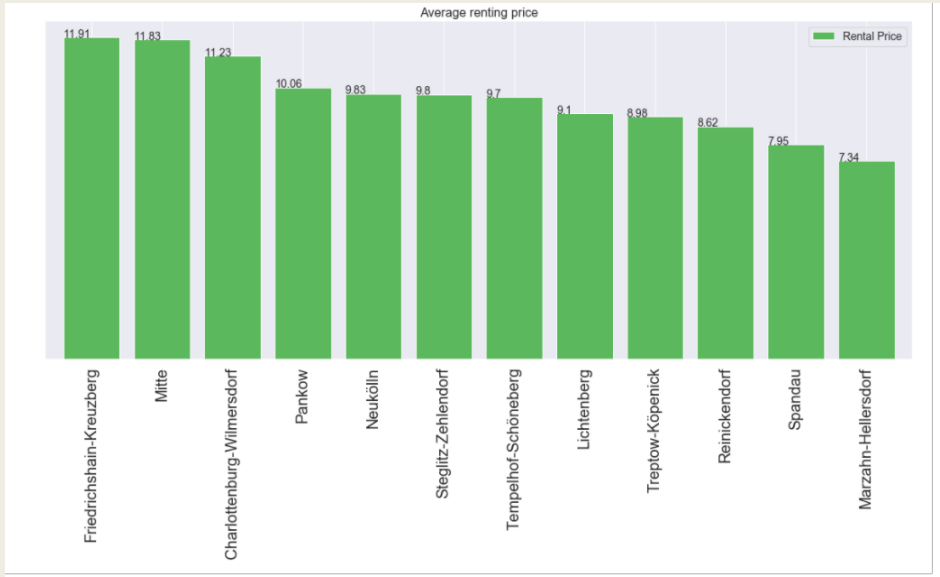
THE CASE OF BERLIN

COURSERA CAPSTONE PROJECT

BACKGROUND & PROBLEM

- Rising rental prices in Berlin City
- But rentring prices in different neighborhoods seem so contradict the proportion of good/middle/ simple residental location(Neighborhoods with high rental prices do not show high propotion of good residental location)





Data: List of Neighborhoods

- List of neighborhoods provided via wikipedia (web scraping)
- Latitude, longitude data via Geopy

	Bezirk	Latitude	Longitude
0	Mitte	52.517690	13.402376
1	Friedrichshain-Kreuzberg	52.515306	13.461612
2	Pankow	52.597663	13.436351
3	Charlottenburg-Wilmersdorf	52.507856	13.263952
4	Spandau	52.519267	13.195439
5	Steglitz-Zehlendorf	52.429205	13.229974
6	Tempelhof-Schöneberg	52.440603	13.373703
7	Neukölln	52.481150	13.435350
8	Treptow-Köpenick	52.417893	13.600185
9	Marzahn-Hellersdorf	52.522523	13.587663
10	Lichtenberg	52.532161	13.511893
11	Reinickendorf	52.604763	13.295287

Data: Statistics of Berlin regarding

- Statistical Data downloaded via official statistic institute of Berlin (export CSV.file)

	Bezirk	Rental Price	Residental area simple	Residental area middle	Residental area good	Population density	No. housing builings	Living space total	Living space per Person	TotalDistanz
0	Mitte	11.83	250873	103743	26485	9368	13230	13242	36.6	0.014169
1	Friedrichshain-Kreuzberg	11.91	154090	131107	2317	13461	9661	10457	38.5	0.074482
2	Pankow	10.06	96619	259768	50051	3833	36993	15743	40.2	0.128118
3	Charlottenburg-Wilmersdorf	11.23	11784	102113	227148	4895	17928	14518	46.2	0.134089
4	Spandau	7.95	151394	86980	3595	2584	27963	8820	37.5	0.195652
5	Steglitz-Zehlendorf	9.80	17560	139351	150800	2851	40621	13441	46.3	0.246717
6	Tempelhof-Schöneberg	9.70	97276	179700	74159	6420	28586	13604	40.2	0.091590
7	Neukölln	9.83	278448	51777	0	7122	27982	11411	35.9	0.082377
8	Treptow-Köpenick	8.98	56411	195542	16217	1561	37405	10278	40.0	0.310468
9	Marzahn-Hellersdorf	7.34	106182	154723	6450	4250	31488	9751	37.7	0.204289
10	Lichtenberg	9.10	66311	210411	12171	5446	17123	10079	35.9	0.138157
11	Reinickendorf	8.62	138833	93281	32466	2897	35701	10137	39.4	0.181299

Data: Foursquare Data

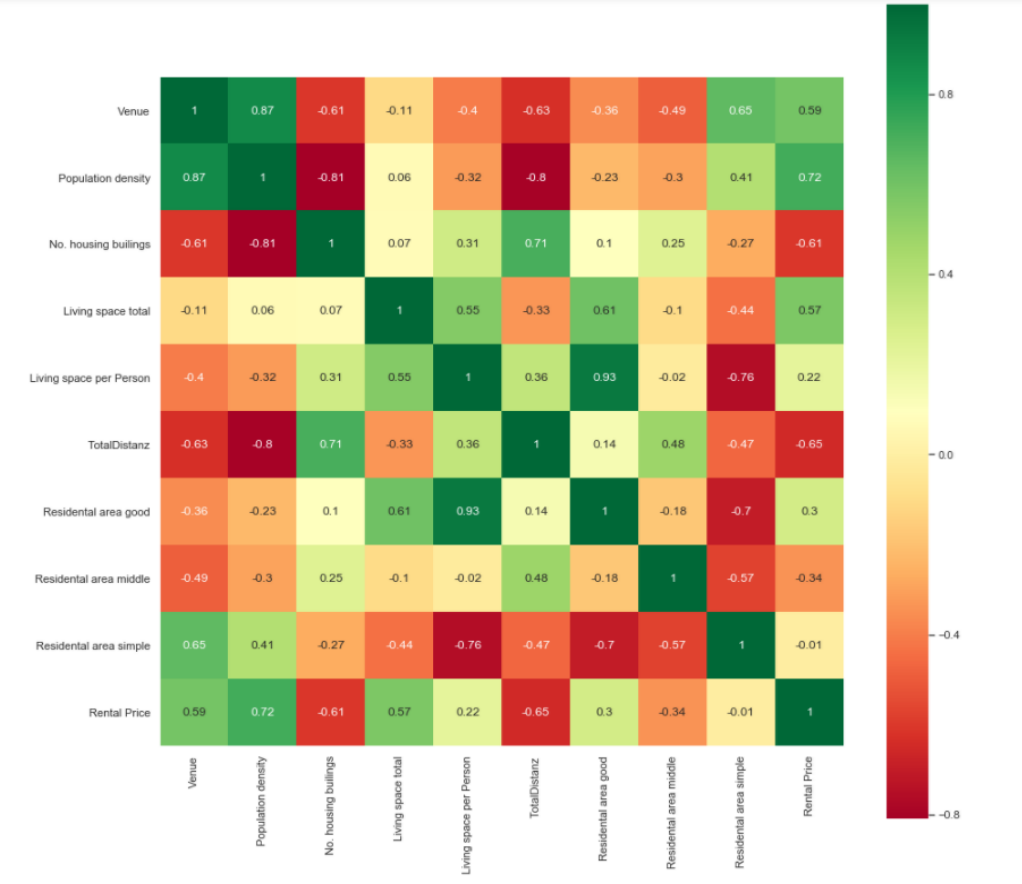
- Foursqure Data:

The table shows that the number of venues per district is reflected in the rents and thus the assumption can be made that there is a connection in this respect

	Bezirk	Venue
0	Charlottenburg-Wilmersdorf	9
1	Friedrichshain-Kreuzberg	100
2	Lichtenberg	8
3	Marzahn-Hellersdorf	12
4	Mitte	59
5	Neukölln	81
6	Pankow	4
7	Reinickendorf	4
8	Spandau	9
9	Steglitz-Zehlendorf	7
10	Tempelhof-Schöneberg	8
11	Treptow-Köpenick	4

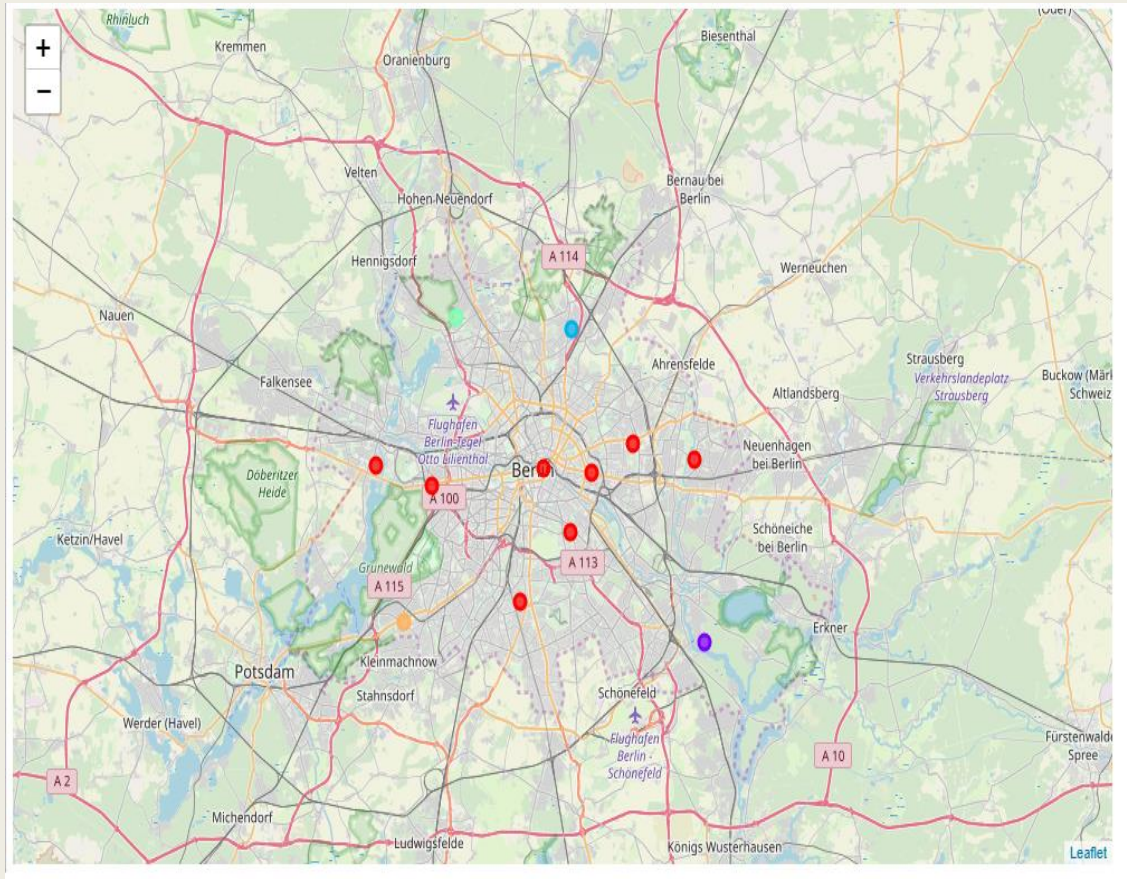
Results: Regression/Correlation

- Correlation table shows, which potential explanatory should be included in the regression model
  - *No. Of Venues*
  - *Population density*
  - *No. Of housing Buildings*
  - *Distance to city center*



Results: Cluster Analysis

- The cluster analysis teaches us how similiar dsitricts can be grouped based on different features. Nevertheless shows the clusteranalysis the typical divide of a city and confirms therefore what was to be expected. Central areas are more similar than districts that are more far ways



Conclusion

- Two approaches:
  - Multiple linear regression (looking which variable has the biggest impact on rental prices)
  - Cluster analyses (cluster similiar neighborhoods)
- Improvements of the model:
  - Check the statistical prerequisites of the data
  - causality
  - more potential explanatory variables